Topics in Response Surface Model Adequacy *Assurance* and *Assessment*

Richard DeLoach
NASA Langley Research Center

The NASA Statistical Engineering Symposium
Williamsburg, VA
May 3-5, 2011

An Alternative Concept of *Quality* in Experimental Aeronautics

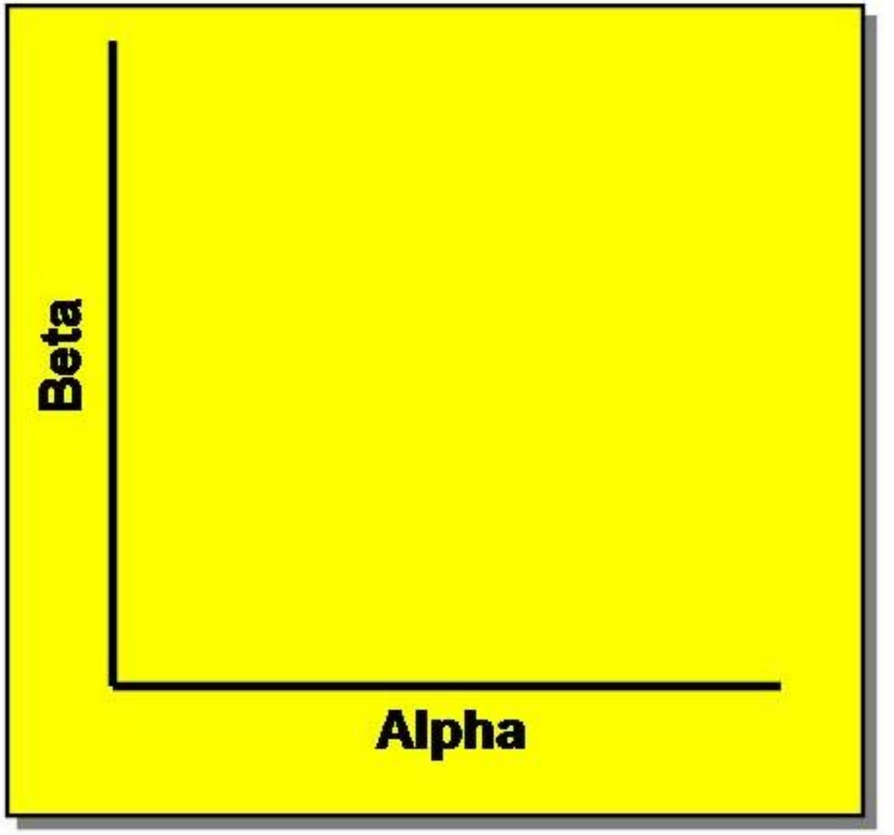
- Traditional concept of quality in wind tunnel testing
  - Data-centric: “Quality” means “Data Quality” in traditional testing
  - Associated with low levels of unexplained variance in a data sample
- An alternative concept of quality
  - Introduced to the Langley experimental aeronautics community in the mid-90’s as the Modern Design of Experiments (MDOE)
  - Associated with inference error probability
  - “Quality” means “getting the right answer”
    - Low probability of inference error
    - Independent of quality of the data

Response Surface Modeling

- **Response Surface Models are mathematical functions representing responses (forces/moments, etc.) as a function of independent variables (AoA, Mach No., etc.)**
- **Quality is cast in terms of modeling adequacy**
  - **For an adequate model, no more than a specified percentage of response predictions are outside acceptable tolerance limits**
  - **Quality, or model adequacy, must be both *assured* and *assessed***
- **Model adequacy is *assured* through the design by**
  - **Data volume specification (*How many* points)**
  - **Site selection within the design space (*Which* points)**
  - **Number and selection of points to be replicated**
  - **Order in which the points are acquired**
- **Model adequacy is *assessed* by examination of residuals**

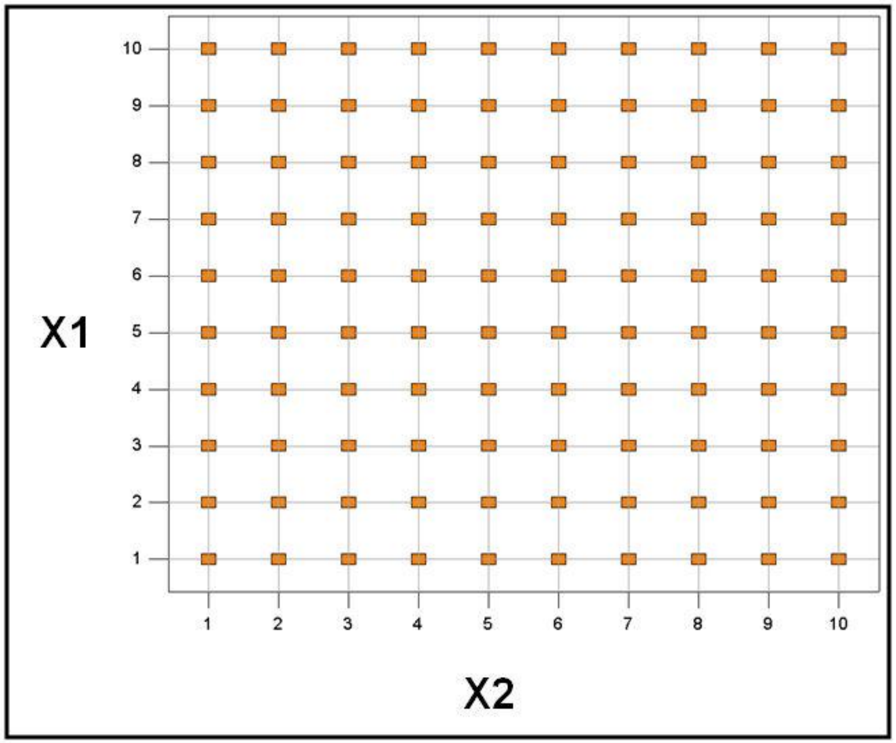
An Inference Space

- A Coordinate System
- One axis for each variable
- Each point represents a unique combination of variable levels
- A response surface is constructed “over” an inference space

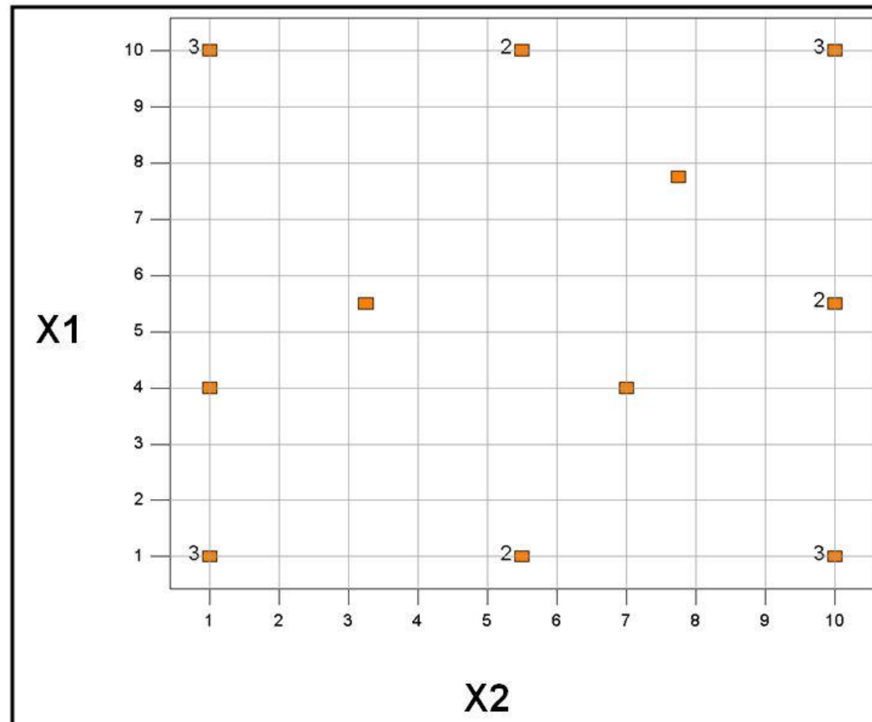


Design Space Comparisons

OFAT

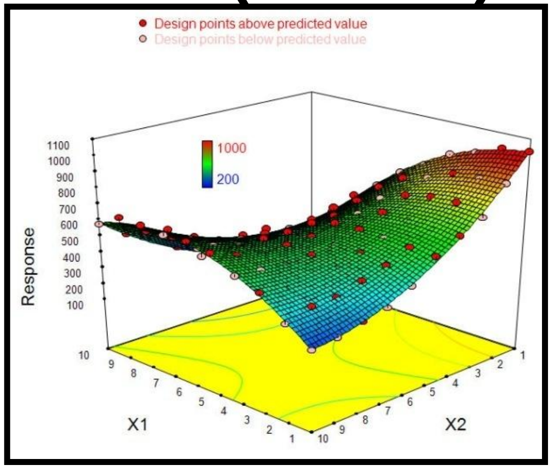


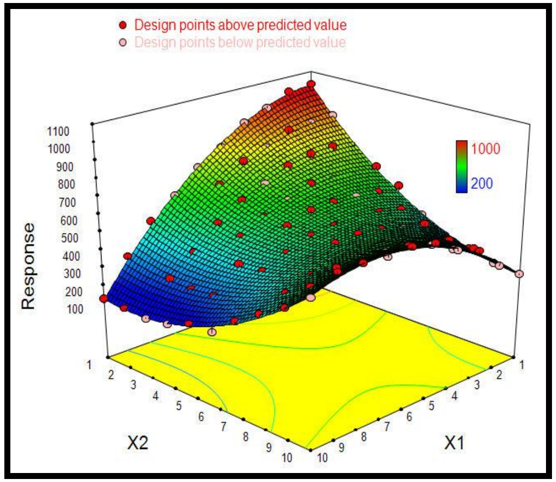
MDOE



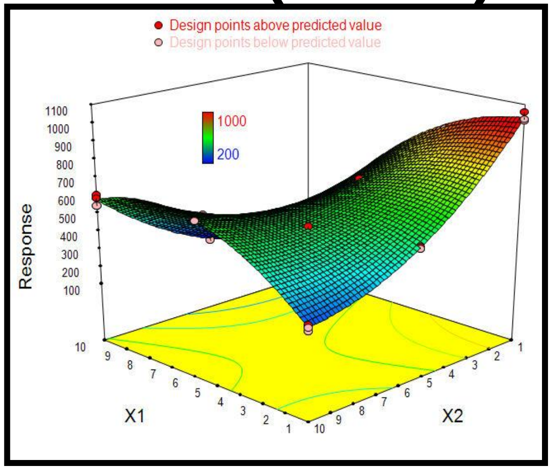
OFAT and MDOE Response Surfaces

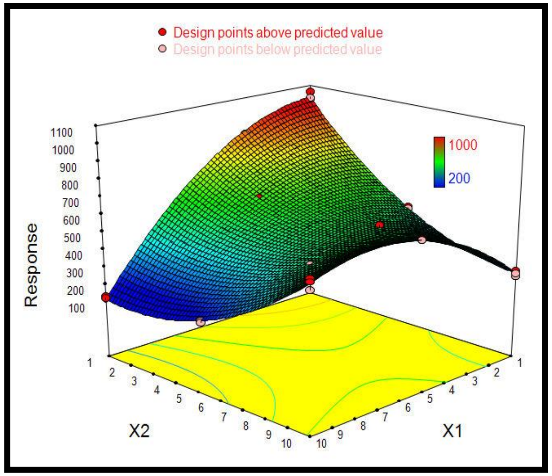
OFAT (100 Pts)





MDOE (22 Pts)



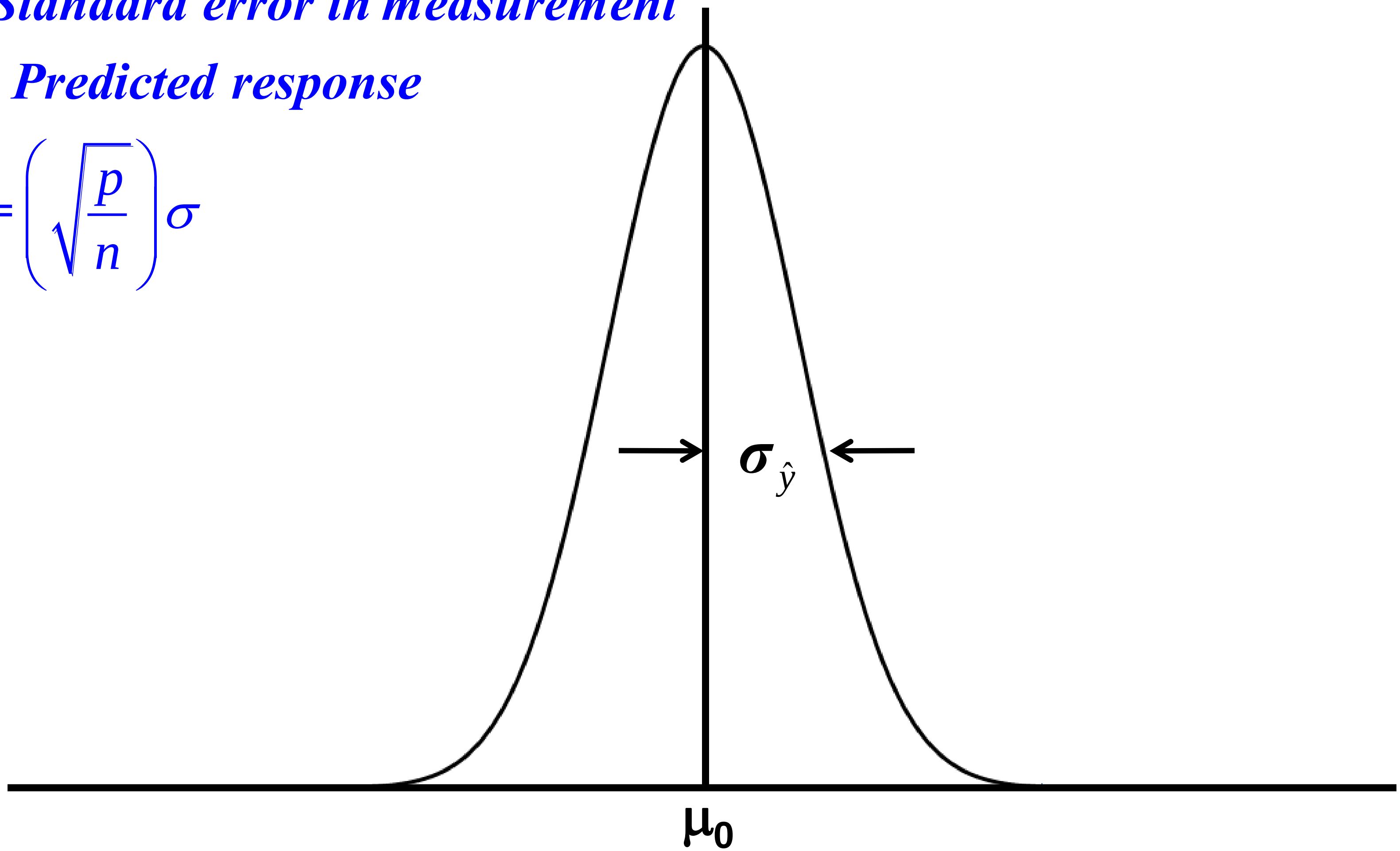


The Mathematics of Quality Assurance and Quality Assessment

σ : Standard error in measurement

μ_0 : Predicted response

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$



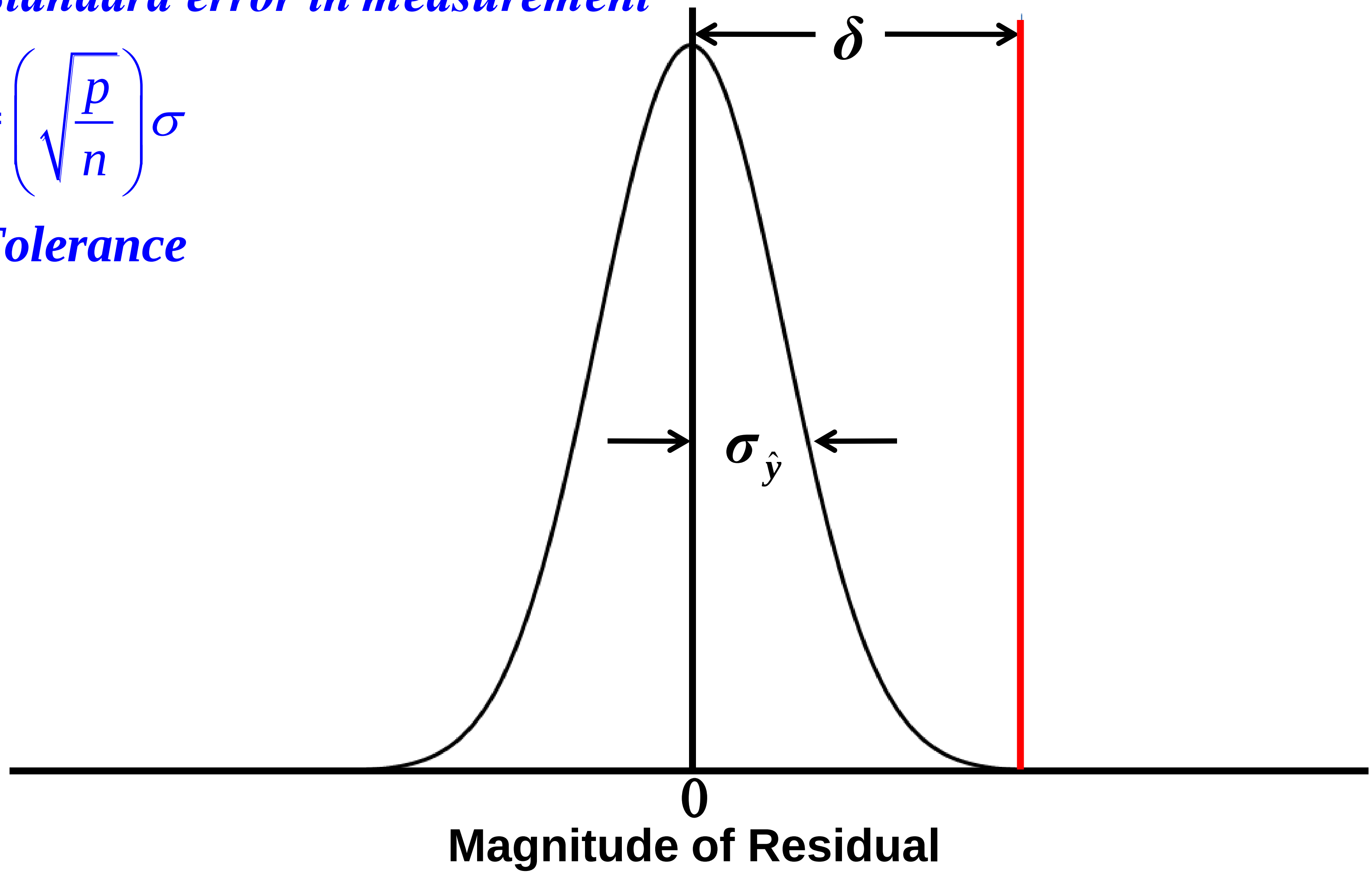
Reference Distribution Under H_0

H_0: Null hypothesis that there is no difference between predicted and measured response (Residual is 0)

σ : *Standard error in measurement*

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$

δ : *Tolerance*



Reference Distributions for Residuals

Black – H_0: True residual is zero

Red – H_A: True residual is borderline unacceptable

σ : *Standard error in measurement*

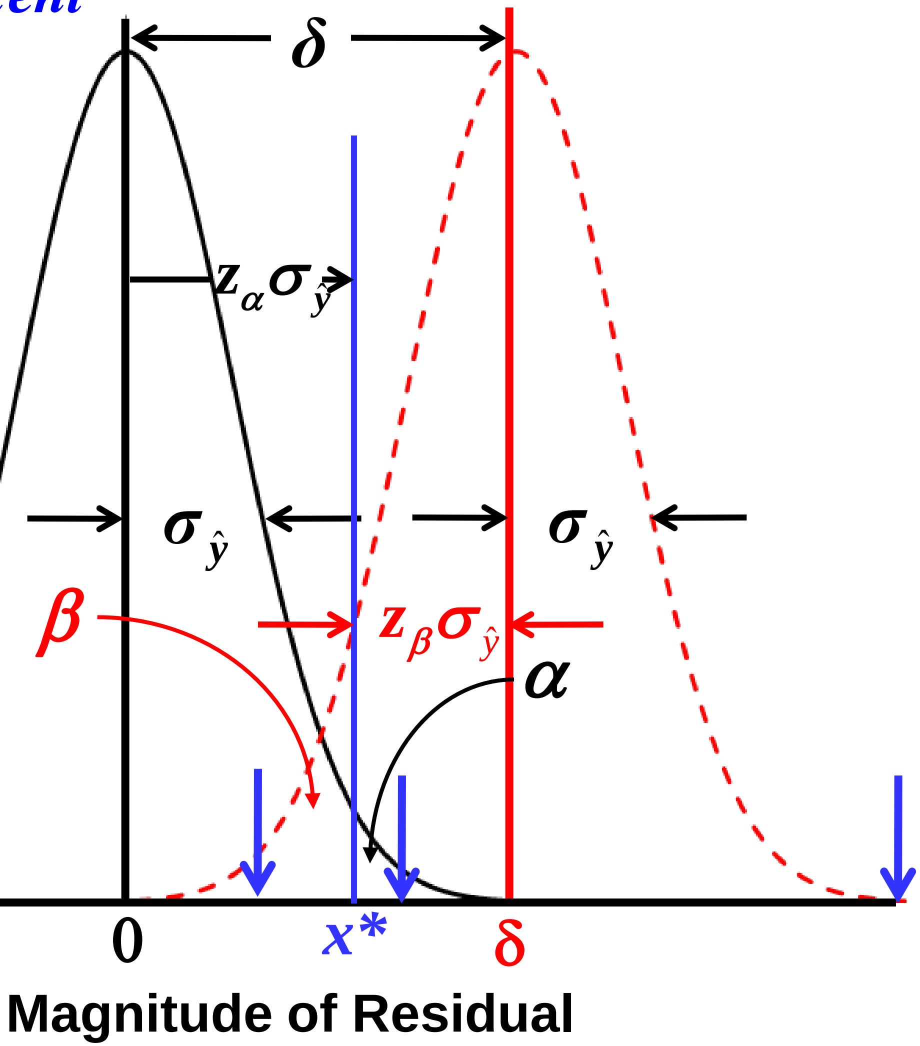
$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$

δ : *Tolerance*

x^* : *Criterion*

α is probability of rejecting adequate model

β is probability of validating erroneous model



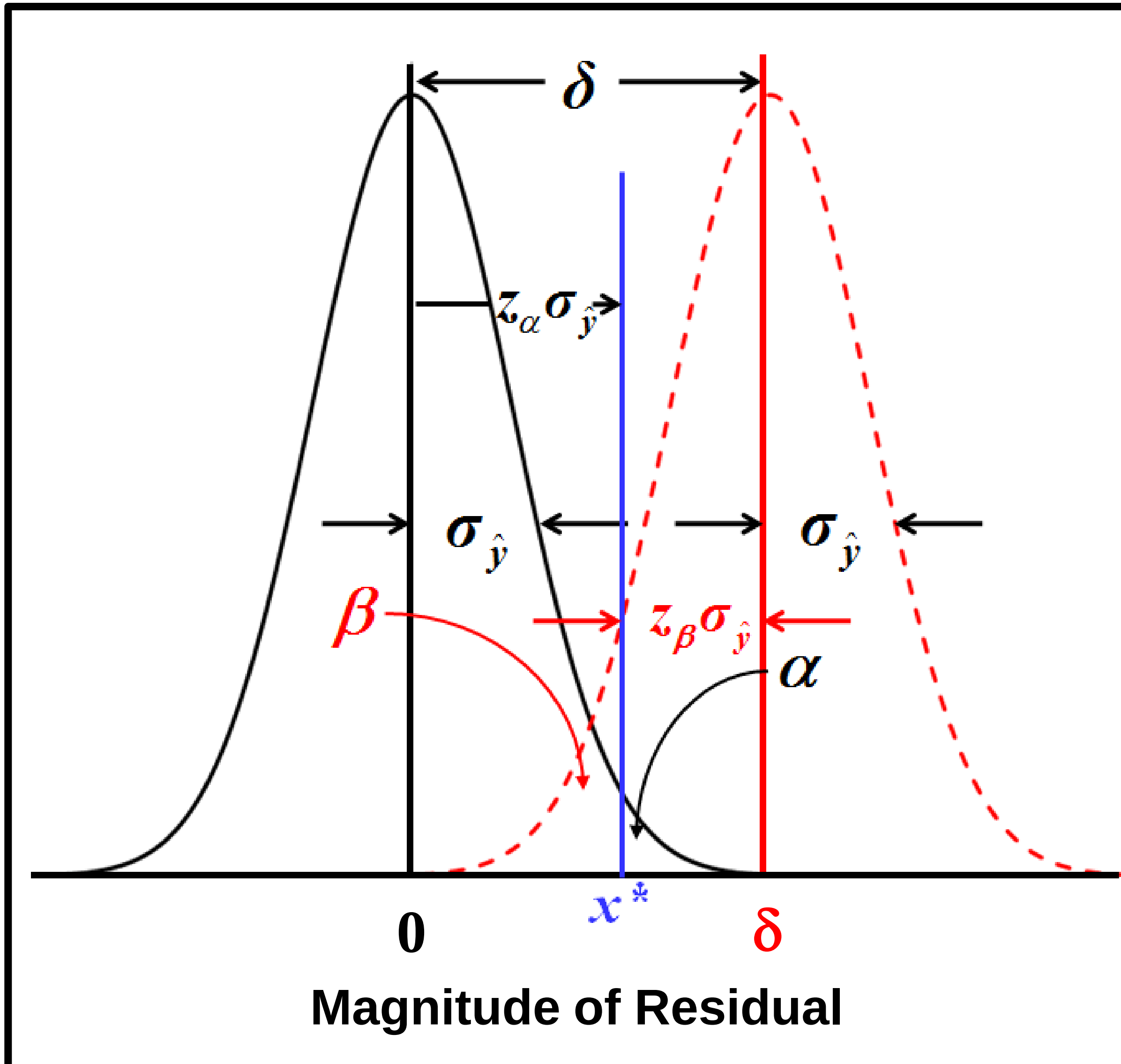
Data Volume Requirement

$$\delta = \left(z_\alpha + z_\beta\right)\sigma_{\hat{y}}$$

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}}\right)\sigma$$

$$\delta^2 = \frac{p\left(z_\alpha + z_\beta\right)^2 \sigma^2}{n}$$

$$\boxed{n = p\left[\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}\right]}$$



Data Volume Formula

Some Practical Difficulties

$$n = p\left[\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}\right]$$

- **The data volume formula depends on five quantities**
  - **Three (p, α, and β) can often be specified by the design consultant**
  - **The tolerance, δ, should be specified by the customer**
  - **The standard measurement error, σ, should be specified by the facility**
- **The customer often prefers to specify tolerance as a multiple of σ, rather than in absolute terms**
  - **A customer may feel comfortable saying his tolerance is "2σ"**
  - **He doesn't always feel he has to know what "σ" is to say this**

Incorporating Tolerance in Data Volume Estimates

- **Consider the general case, in which $\delta = K\sigma$, where *K* is a constant specified by the customer**
- **Note that a specific "*K*" may eventually evolve as an industry convention (about which more in a moment)**

$$\delta = K\sigma$$

$$n = p\left(z_\alpha + z_\beta\right)^2 \frac{\sigma^2}{\delta^2}$$

$$n = \left(\frac{z_\alpha + z_\beta}{K}\right)^2 p$$

Special Case for Tolerance, δ

- **We have for the general case in which $\delta = K\sigma$:**

$$n = \left(\frac{z_\alpha + z_\beta}{K} \right)^2 p$$

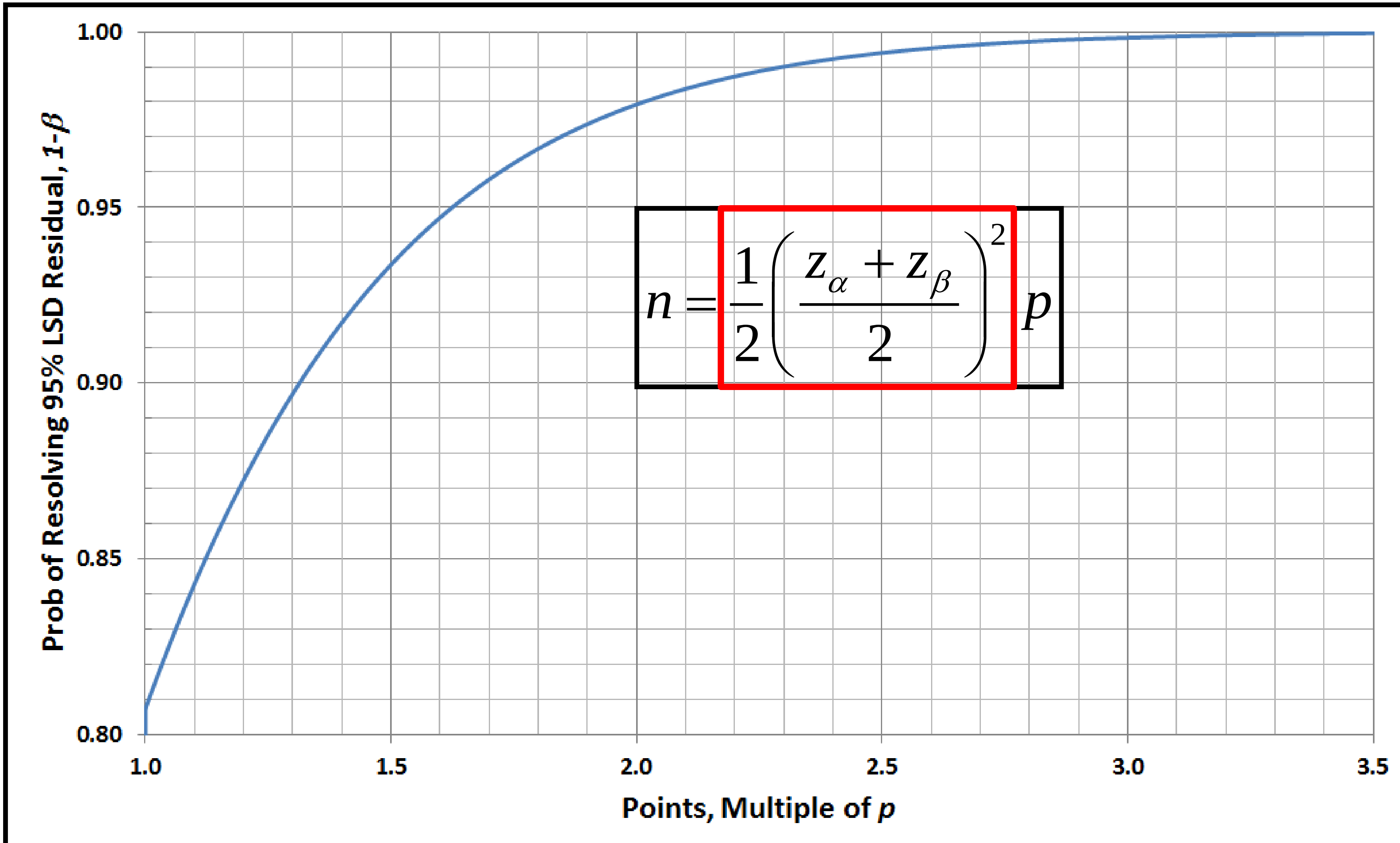
- **Let δ = 95% LSD (Least Significant Difference)**
  - **This is the smallest difference between two replicated measurements that can be resolved with 95% confidence**
  - **It may be regarded as a reasonable tolerance specification**

$$\delta = 95\%\ \mathrm{LSD} = 2\sqrt{2}\sigma \rightarrow K = 2\sqrt{2}$$

$$n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2} \right)^2 p$$

Model Term-Count Multiplier

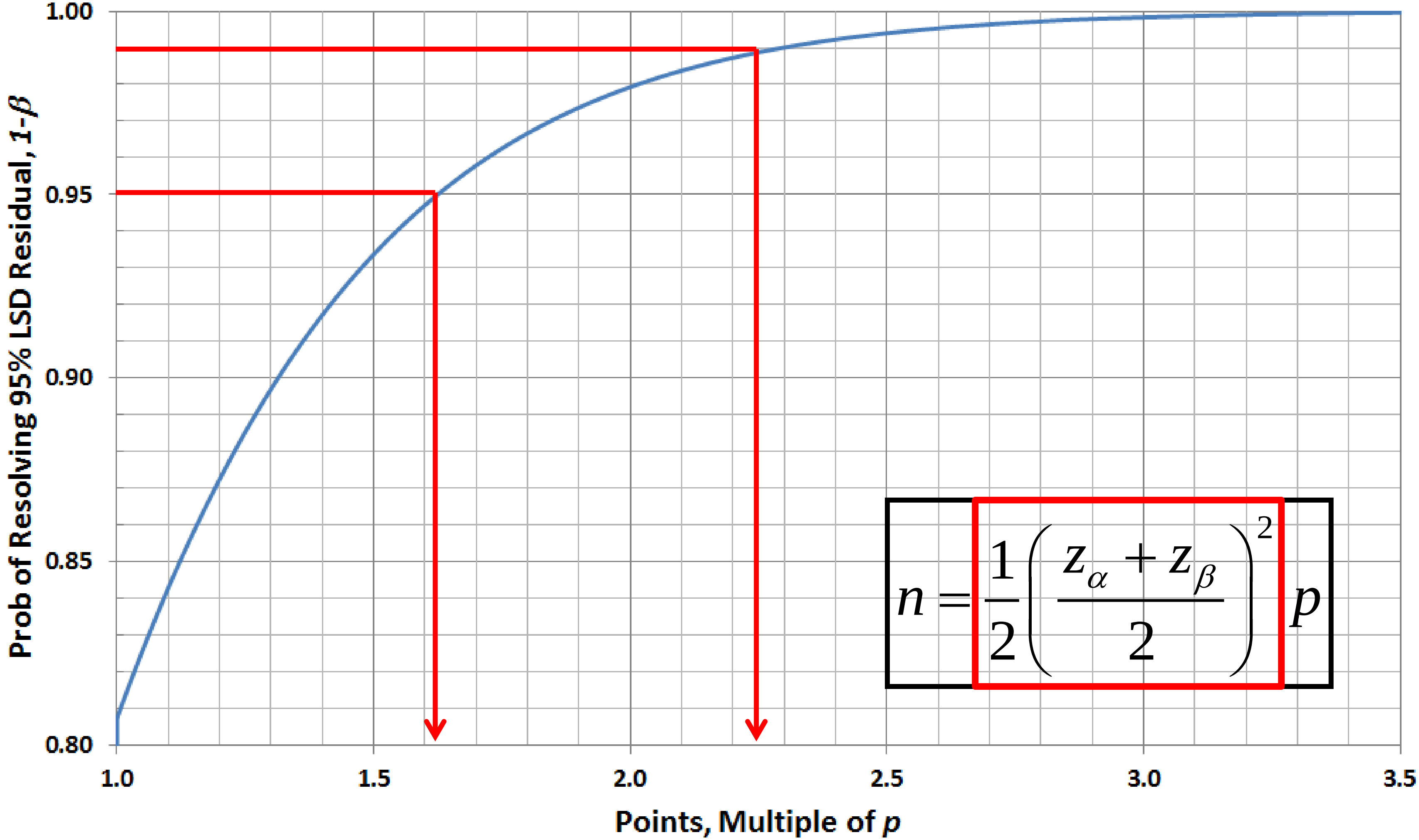
Minimum to Resolve 95% LSD with $\alpha = 0.05$



$$n = \boxed{\frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2} p$$

Model Term-Count Multiplier

Minimum to Resolve 95% LSD with $\alpha = 0.05$



$$n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2 p$$

Langley Research Center

Another Special Case for Tolerance, δ

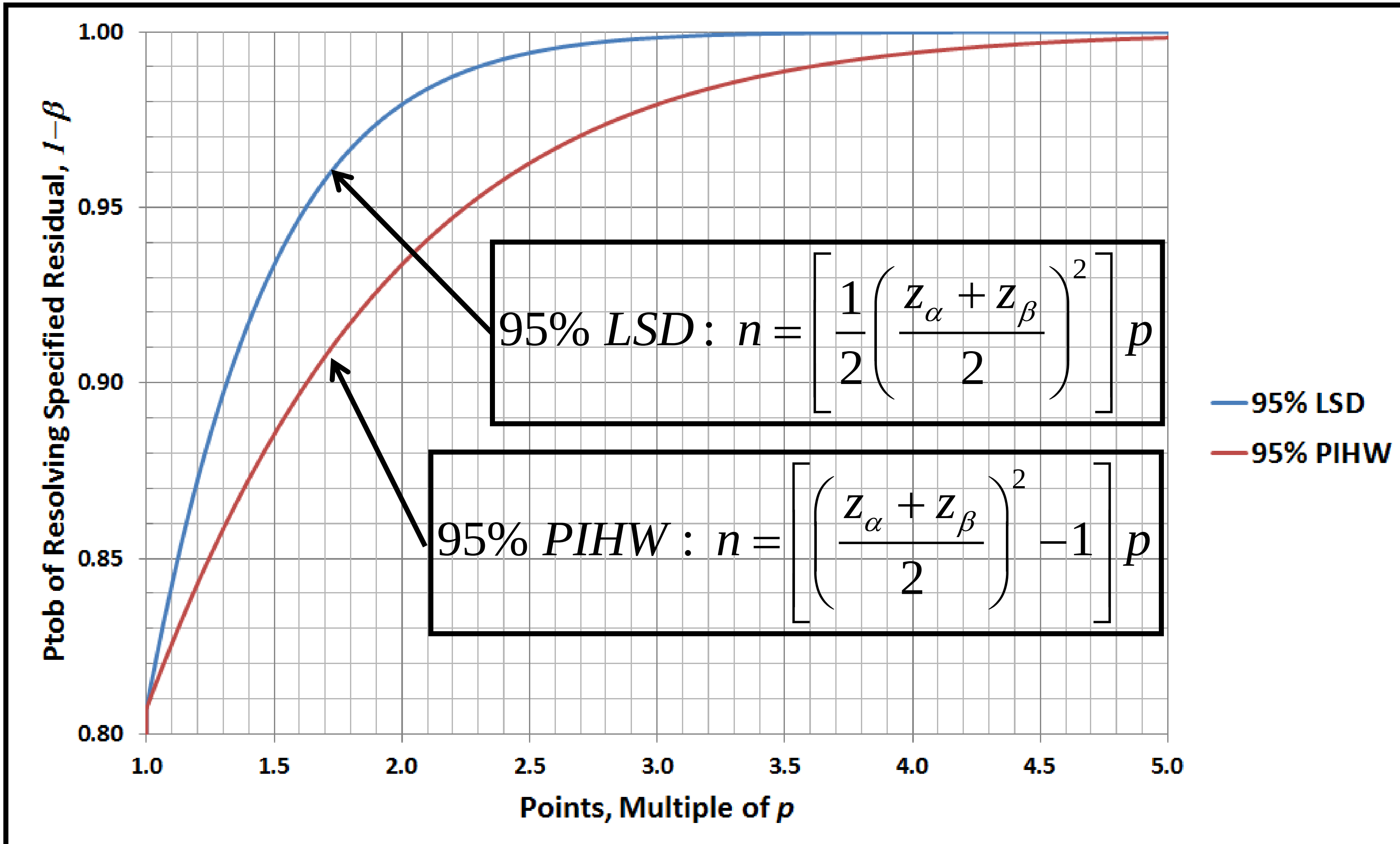
- **Let δ = 95% PIHW (Prediction Interval Half-Width)**
  - **This is the smallest difference between a physical measurement and a *model prediction* that can be resolved with 95% confidence**
  - **It is a convenient tolerance spec because most curve-fitting software packages compute this automatically**

$$95\%\ \text{PIHW} = 2\sqrt{\left(1+\frac{p}{n}\right)}\sigma \rightarrow K = 2\sqrt{\left(1+\frac{p}{n}\right)}$$

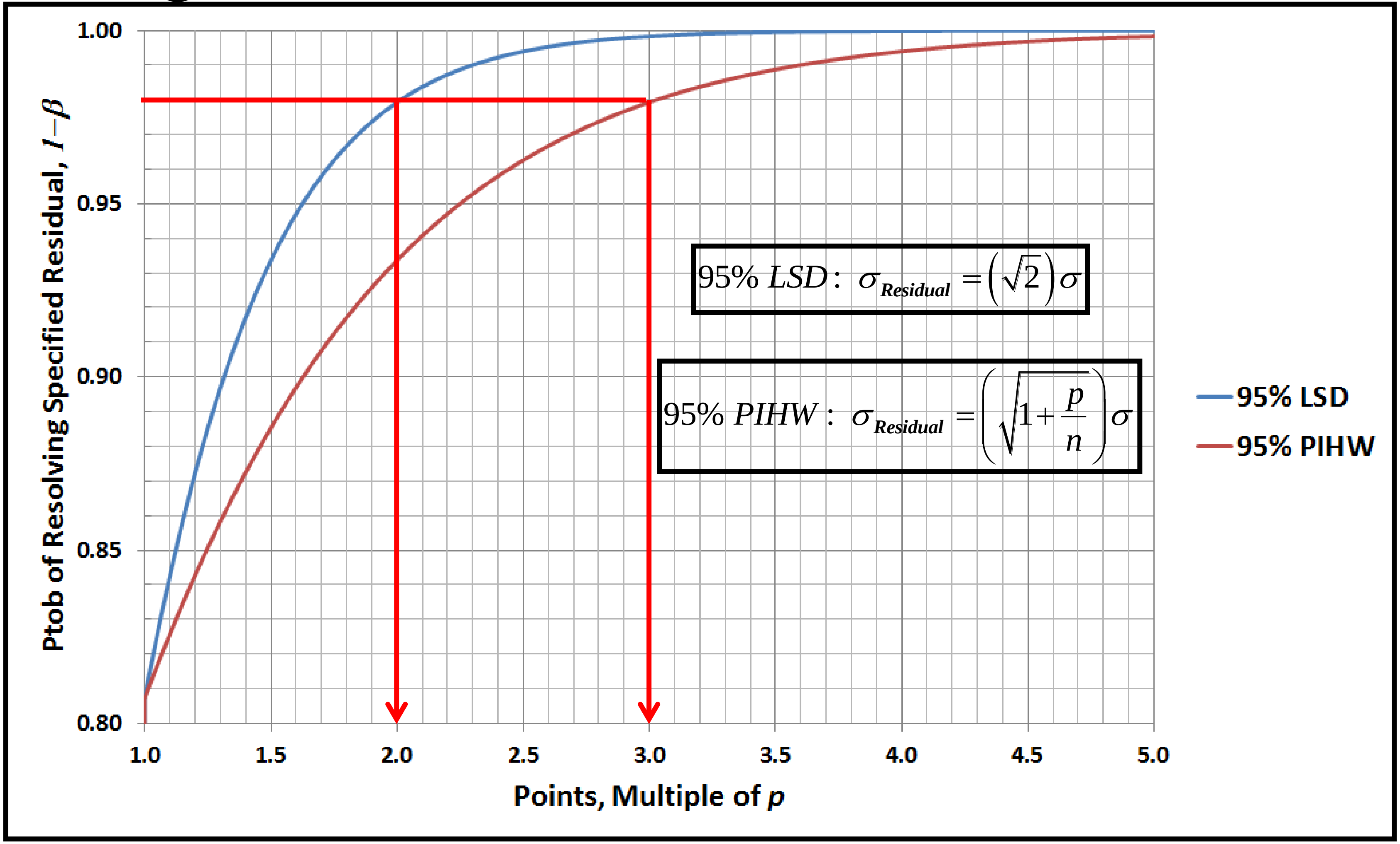
$$n = \left[\boxed{\left(\frac{z_\alpha + z_\beta}{2}\right)^2} - 1\right]p$$

Model Term-Count Multiplier

Minimum to Resolve 95% LSD or 95% PIHW with $\alpha = 0.05$



95% PIHW Tolerance Criterion is More Stringent than the 95% LSD Criterion



Numerical Scaling Example

Typical OFAT Wind Tunnel Test

- **Consider a wind tunnel test in which forces and moments are to be estimated as a function of four factors**
  - **Angles of Attack and Sideslip**
  - **Mach Number**
  - **Height (for ground effects)**
- **Typical OFAT levels might be as follows**
  - **AoA: -5^o to +15^o in 1^o increments (21 levels)**
  - **Sideslip: 0^o to +10^o in 2^o increments (6 levels)**
  - **Mach Number from 0.70 to 0.90 in 0.25 increments (9 levels)**
  - **Height (5 levels)**
- **Total of 21 x 6 x 9 x 5 = 5670 points (not atypical for OFAT test)**
- **Standard error in response estimate: σ**

Numerical Scaling Example

Corresponding MDOE Scaling Case

- **Assume adequate fits can be achieved over three AoA sub-ranges and two sideslip sub-ranges with 4th-order models**
- **A *d*th-order model in *k* factors has *p* terms (including intercept), where**

$$p = \frac{(d+k)!}{d!k!} = \frac{(4+4)!}{4!4!} = 70$$

- **Assume a 95% LSD tolerance specification:**

$$\boxed{n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2 p}$$

Numerical Scaling Example, Cont.

- **Specify inference error risk tolerances**
  - **Max acceptable probability of rejecting a valid model: $\alpha = 0.05$**
  - **Max acceptable probability of validating a bad model: $\beta = 0.01$**
- **Look up corresponding standard normal deviates, $z_{\alpha,\beta}$**
  - **For $\alpha = 0.05$, $z_\alpha = 1.960$ (double-sided null hypothesis)**
  - **For $\beta = 0.01$, $z_\beta = 2.326$ (single-sided alternative hypothesis)**
- **Estimate data volume per subspace:**

$$n = \frac{1}{2}\left(\frac{z_\alpha + z_\beta}{2}\right)^2 p = \frac{1}{2}\left(\frac{1.960 + 2.326}{2}\right)^2 70 = 2.296\,p = 161$$

- **Estimate total data volume (six subspaces): 6 x 161 = 966**

MDOE/OFAT Comparison

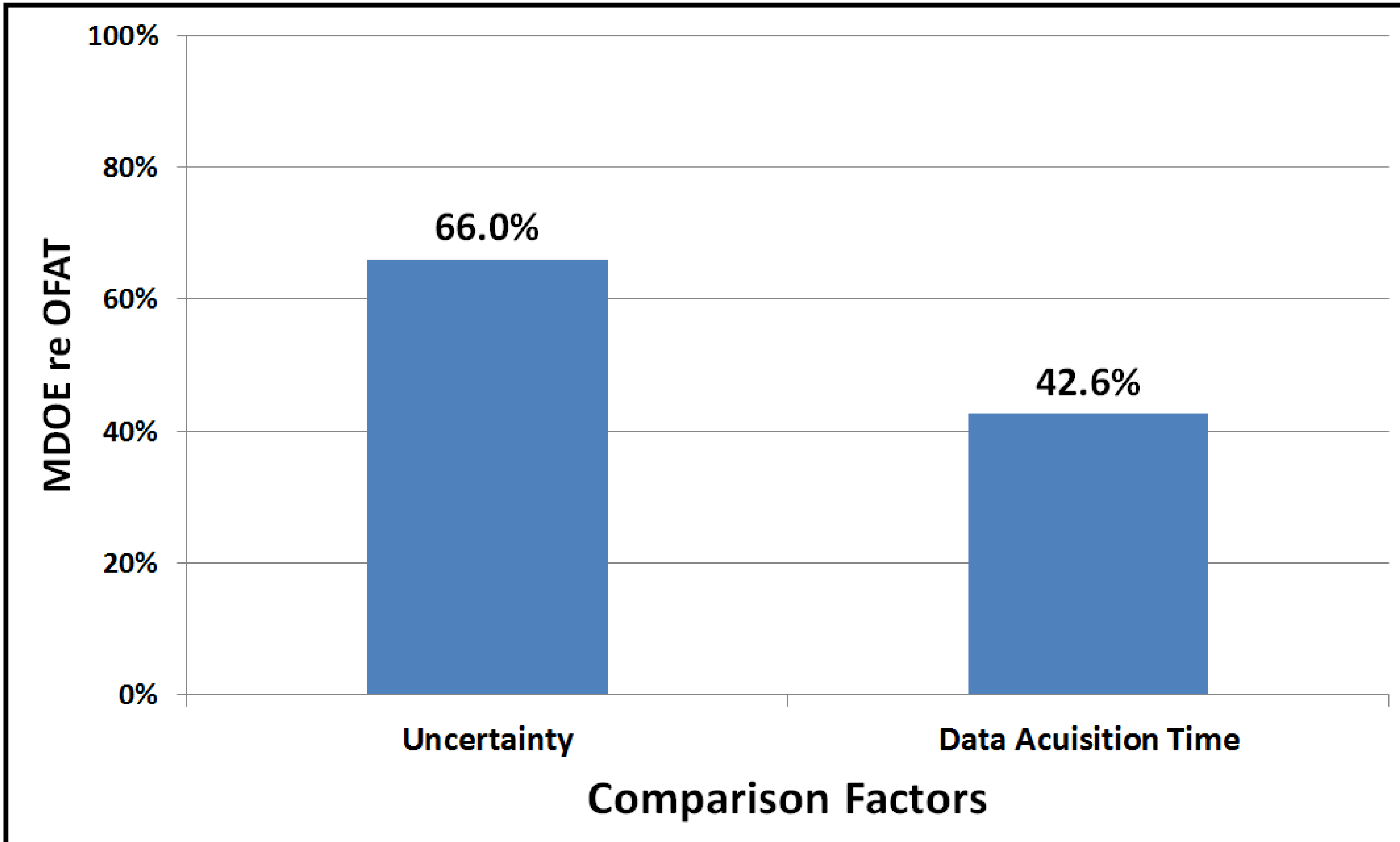
- **There is a large apparent difference in OFAT and MDOE resource requirements**
  - **OFAT: 5670 points**
  - **MDOE: 966 points**
- **The savings are not that dramatic, however**
- **MDOE methods invoke certain quality assurance tactics to defend against *covariate effects***
- **Covariates are slowly varying, persisting factors that are not controlled by the experimenter**
  - **They are generally larger than ordinary random variations**
  - **They are not reproducible from test to test**
  - **They are largely overlooked in conventional OFAT testing**

MDOE/OFAT Comparison, Cont.

- **MDOE quality assurance tactics to defend against covariates cost a factor of 1.5 to 2.5 in data rate**
  - **In the time needed to acquire 966 MDOE points, up to 2.5 times as many OFAT points might be acquired**
  - **This would be 2.5 x 966 = 2415 points**
- **The MDOE data acquisition time is thus expected to be no more than a factor of 2415/5670 of the OFAT requirement, or 42.6% (and could be rather less)**
- **Note that the scaling resulted in a data volume requirement of *n = 2.296p***
- **The MDOE standard error is thus**

$$\sigma_{\hat{y}} = \left(\sqrt{\frac{p}{n}} \right) \sigma = \left(\sqrt{\frac{p}{2.296p}} \right) \sigma = 0.660\sigma$$

Quality and Productivity Comparison



Concluding Remarks

- **Quality in wind tunnel testing is more properly expressed in terms of inference error probability than unexplained variance in the raw data**
  - **It is more important to get the *right answer*, than “good data”**
  - **This imposes a responsibility to articulate tolerance requirements**
- **There is a mathematical relationship between resource requirements and quality requirements**
- **Each new data point reduces inference error risk**
  - **Too little data means unacceptable inference error risk**
  - **Too much data means wasted resources**
- **The experimental aeronautics community might consider adopting the 95% LSD as a tolerance specification**
- **Then data volume in the range of 2 to 3 times the number of points needed to fit a model would typically be sufficient**

NASA